AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

BIOTECHNOLOGY

Comparison of Genome and Plasmid-Based Engineering of Multigene Benzylglucosinolate Pathway in *Saccharomyces cerevisiae*

Cuiwei Wang,[a] Michal Poborsky,[a] Christoph Crocoll,[a] Christina Spuur Nødvig,[b]* Uffe Hasbro Mortensen,[b] Barbara Ann Halkier[a]

[a]DynaMo Center, Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg, Denmark
[b]Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark

ABSTRACT Intake of brassicaceous vegetables such as cabbage is associated with numerous health benefits. The major defense compounds in the Brassicales order are the amino acid-derived glucosinolates that have been associated with the health-promoting effects. This has primed a desire to build glucosinolate-producing microbial cell factories as a stable and reliable source. Here, we established—for the first time—production of the phenylalanine-derived benzylglucosinolate (BGLS) in *Saccharomyces cerevisiae* using two different engineering strategies: stable genome integration versus plasmid-based introduction of the biosynthetic genes. Although the plasmid-engineered strain showed a tendency to generate higher expression level of each gene (except *CYP83B1*) in the biosynthetic pathway, the genome-engineered strain produced 8.4-fold higher BGLS yield compared to the plasmid-engineered strain. Additionally, we optimized the genome-engineered strain by overexpressing the entry point genes *CYP79A2* and *CYP83B1*, resulting in a 2-fold increase in BGLS production but also a 4.8-fold increase in the level of the last intermediate desulfo-benzylglucosinolate (dsBGLS). We applied several approaches to alleviate the metabolic bottleneck in the step where dsBGLS is converted to BGLS by sulfotransferase, SOT16 dependent on 3′-phosphoadenosine-5′-phosphosulfate (PAPS). BGLS production increased 1.7-fold by overexpressing *SOT16* and 1.7-fold by introducing APS kinase, APK1, from *Arabidopsis thaliana* involved in the PAPS regeneration cycle. Modulating the endogenous sulfur assimilatory pathway through overexpression of *MET3* and *MET14* resulted in 2.4-fold to 12.81 μmol/L (=5.2 mg/L) for BGLS production.

IMPORTANCE Intake of brassicaceous vegetables such as cabbage is associated with numerous health benefits. The major defense compounds in the Brassicales order are the amino acid-derived glucosinolates that have been associated with the health-promoting effects. This has primed a desire to build glucosinolate-producing microbial cell factories as a stable and reliable source. In this study, we engineered for the first time the production of phenylalanine-derived benzylglucosinolate in *Saccharomyces cerevisiae* with two engineering strategies: stable genome integration versus plasmid-based introduction of the biosynthetic genes. Although the plasmid-engineered strain generally showed higher expression level of each gene (except *CYP83B1*) in the biosynthetic pathway, the genome-engineered strain produced higher production level of benzylglucosinolate. Based on the genome-engineered strain, the benzylglucosinolate level was improved by optimization. Our study compared different approaches to engineer a multigene pathway for production of the plant natural product benzylglucosinolate. This may provide potential application in industrial biotechnology.

KEYWORDS glucosinolate, genome integration, targeted proteomics, *Saccharomyces cerevisiae*, gene copy number

Editor Jeremy D. Semrau, University of Michigan-Ann Arbor

Address correspondence to Barbara Ann Halkier, bah@plen.ku.dk.

*Present address: Christina Spuur Nødvig, Novozymes, Kongens Lyngby, Denmark.

The authors declare no conflict of interest.

Received 16 June 2022
Accepted 9 October 2022
Published 3 November 2022

Glucosinolates (GLSs) are amino acid-derived defense compounds characteristic of the Brassicales order, including vegetables like broccoli and cabbages, as well as the model plant *Arabidopsis thaliana* (1). Intake of GLSs via consumption of brassicaceous vegetables has been associated with numerous health benefits as well as prebiotic effects (2). This has primed a desire to engineer the production of health-promoting GLSs as dietary supplements to enable intake of well-defined doses from a stable and rich source. Multiple biotechnological approaches using either plants or microbes as host organisms have been applied to engineer the multigene pathways for production of the desirable GLSs (3 to 5).

The first example of heterologous production of GLS was in 2009 when benzylglucosinolate (BGLS) was produced by transient expression of the BGLS biosynthetic genes in *Nicotiana benthamiana* (6). In 2012, a versatile platform for stable expression of multigene pathways in the *Saccharomyces cerevisiae* was developed, and 14 locations in the yeast genome were identified for gene integration with high gene expression and without any significant impact on the growth rate (7). As a proof of concept, production of indole GLS in a microbial host was successfully obtained for the first time using this platform (7). In 2019, BGLS was produced from genes on plasmids in *Eschericia coli* (3), and in 2020 4-methylsulfinylbutyl GLS was produced by genome integration of the genes into *E. coli* (8). In addition to using either plasmid or genome engineering as illustrated here for GLS production, recent examples of engineering multigene pathways of alkaloids successfully apply a combination of genome integration and plasmids (9 to 12).

A critical part of engineering multigene pathways is to be able to monitor the expression of the pathway proteins. Compared to Western blot analysis, which is not easily applicable to multiple enzyme pathways as many antibodies would be required, targeted proteomics has emerged as an attractive alternative. The relative protein levels are quantified by monitoring proteotypic peptides of the biosynthetic enzymes using mass spectrometry (13). Recently, targeted proteomics was used to evaluate an optimization strategy for production of BGLS in *E. coli* by monitoring the role of different promoters at the enzyme level (3).

In this study, we investigated multicopy plasmid-based versus genome engineering strategies for BGLS production in *S. cerevisiae*. The single-copy genome integration approach resulted in a higher yield of BGLS than multicopy plasmid expression despite more abundant overall protein level in BGLS biosynthesis by plasmid expression. Additionally, genome-engineered BGLS production was further optimized by overexpression of the entry point enzymes, and strategies for alleviating the metabolic bottleneck in the sulfotransferase step were tested by either overexpressing the responsible sulfotransferase enzyme or modulating related sulfur metabolism.

RESULTS

Comparison of BGLS production and protein level of biosynthetic enzymes in $BGLS_g$ and $BGLS_p$ strains of *S. cerevisiae*. To achieve BGLS production in *S. cerevisiae*, seven biosynthetic genes (*CYP79A2*, *CYP83B1*, *GSTF9*, *GGP1*, *SUR1*, *UGT74B1*, and *SOT16*; see Fig. 1), as well as the gene *ATR1* encoding a NADPH cytochrome P450 reductase that supports the function of cytochrome P450 enzymes from *A. thaliana*, were introduced into the *S. cerevisiae* strain CEN.PK 113-11C. Specifically, the eight genes driven by the strong constitutive promoters TEF1 or PGK1 were pairwise integrated into four well-characterized sites on the chromosome XII (7), and this engineered strain was named $BGLS_g$ (Fig. 2A). For plasmid-based production, two 2μ-derived high-copy plasmids (containing *URA3* and *HIS3* as selectable markers, respectively) containing the eight genes under the control of the same promoters and terminators as in the $BGLS_g$ strain were constructed and transformed into CEN.PK 113-11C, and the strain was named $BGLS_p$ (Fig. 2B). As control, we transformed the wild-type strain CEN.PK 113-11C with the two empty plasmids that were used for expression of the BGLS pathway genes in the $BGLS_p$ strain. Additionally, we transformed $BGLS_g$ with the two empty plasmids for comparison with the strain $BGLS_p$ in the same growth medium. When the

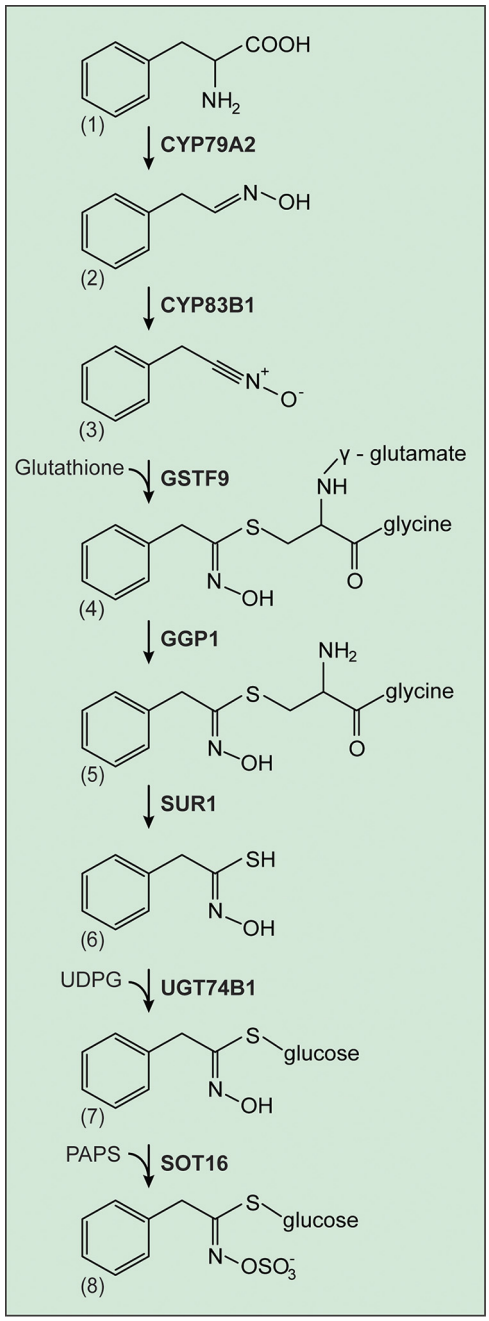

FIG 1 The BGLS biosynthetic pathway in *Arabidopsis thaliana*. The compounds in the BGLS pathway are (i) phenylalanine, (ii) phenylacetaldoxime, (iii) phenylacetonitrile oxide, (iv) S-[(Z)-phenylacetohydroximoyl]-L-glutathione, (v) Cys-Gly conjugate derivative of compound iv, (vi) phenylacetothiohydroximic acid, (vii) desulfo-benzylglucosinolate (dsBGLS), and (viii) benzylglucosinolate (BGLS). Abbreviations: GSTF9, glutathione *S*-transferase F9; GGP1, γ-glutamyl peptidase 1; SUR1, C-S lyase; UGT74B1, UDP-glucosyltransferase 74B1; SOT16, sulfotransferase 16; UDPG, uridine-5′-diphosphate-glucose, PAPS, 3′-phosphoadenosine-5′-phosphosulfate; PAP, adenosine-3′,5′-bisphosphate.

three strains were grown in SC-Ura-His-Glu medium for 48 h, we found that both $BGLS_g$ and $BGLS_p$ strains produced BGLS (Fig. 3A), whereas the control strain did not (Table S1 in the supplemental material). The $BGLS_g$ strain produced 0.59 μmol/L BGLS, while the $BGLS_p$ produced 0.07 μmol/L (Table S1). This shows that expression of the BGLS biosynthetic genes through stable genome integration results in 8.4-fold higher production level of BGLS compared to expression from the plasmids despite the much higher copy number of plasmids compared to the single copy integrated into the

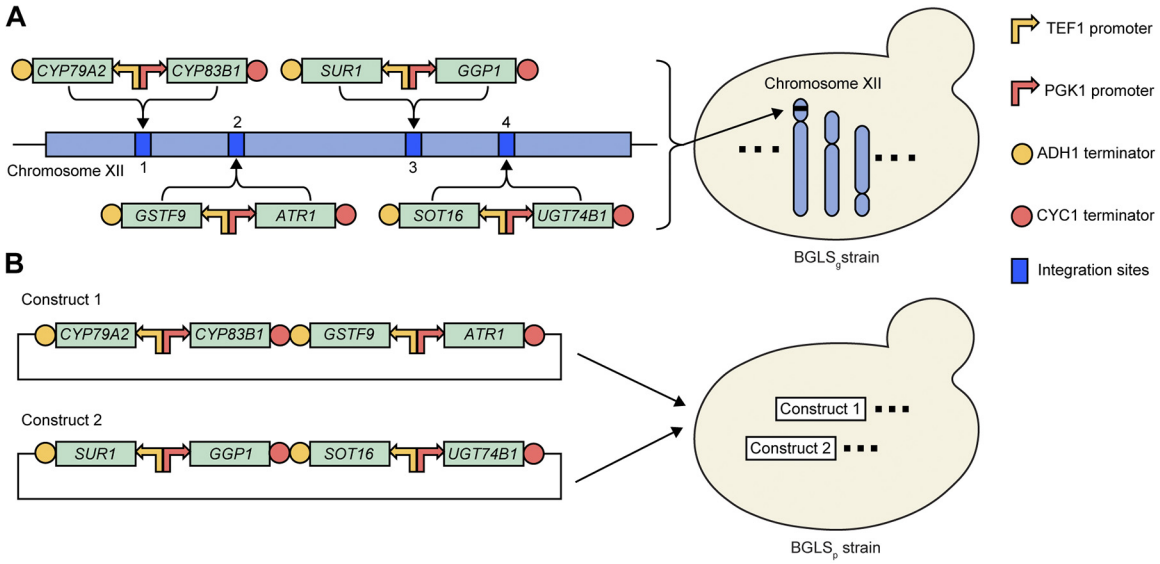

FIG 2 Two strategies for engineering the production of BGLS in the *Saccharomyces cerevisiae* strain. (A) Strategy for stable integration of eight genes of the BGLS pathway into the genome of *S. cerevisiae*. Pairs of two genes flanked with, respectively, TEF1 promoter and ADH1 terminator or PGK1 promoter and CYC1 terminator were inserted into the integration sites at chromosome XII. The strain is named BGLS$_g$. (B) Strategy for engineering the BGLS pathway genes in two plasmids. The genes are flanked with the same promoters and terminators shown in A. *CYP79A2*, *CYP83B1*, *GSTF9*, and *ATR1* were constructed on high-copy-number plasmid pESC-URA-USER, and the remaining four genes are constructed on high-copy-number plasmid pESC-HIS. The strain is named BGLS$_p$. The support gene *ATR1* is coexpressed with the BGLS pathway genes shown in A and B.

genome. Furthermore, we observed that both BGLS$_g$ and BGLS$_p$ produced a small amount of desulfo-benzylglucosinolate (dsBGLS), which is the last intermediate in the BGLS biosynthetic pathway (Fig. 3B). Noticeably, large differences between five biological replicates for BGLS production were observed, as evidenced by the results that one of five colonies did not produce detectable level of BGLS and that a large variation of dsBGLS yields was detected across the individual BGLS$_p$ cultures.

We further analyzed the protein levels of the individual biosynthetic enzymes in the BGLS$_g$ and BGLS$_p$ strains by targeted proteomics when the two strains were grown for 24 h. Generally, we find that the average protein level for most BGLS enzymes shows a tendency to be higher when the genes are expressed from high-copy plasmids (ranging from 0.02 to 1.35) than from single copies integrated into the genome (ranging from 0.02 to 0.34). An exception is CYP83B1, which has a similar protein level in the two strains (Fig. 3C to I and Table S1). It seems that overall variation in protein levels of CYP83B1 reflected the varied BGLS production yields in BGLS$_g$ strain as judged from Fig. 3A and D. However, when looking at the individual strains, the CYP83B1 protein level in each culture was not consistent with the tendency of BGLS production level in the same culture (Fig. S1A and D). Although CYP79A2 was previously detected in BGLS production in *E. coli* (3), the protein was not detectable in our experimental setup. Interestingly, large variation among individual replicates for each enzyme was observed in BGLS$_p$ with high-copy plasmids, which leads to unbalanced enzyme levels within a biosynthetic pathway in a yeast cell. The poor orchestration among the enzymes for BGLS biosynthesis explains the lower BGLS yield despite the higher average of protein level in BGLS$_p$ (Fig. S1). Specifically, one of the cultures (C3) expressing undetectable levels of CYP83B1 and SUR1 did not produce traceable levels of BGLS and dsBGLS, and another culture (C4) with undetectable levels of CYP83B1 produced a tiny amount of BGLS and no dsBGLS. Moreover, the undetectable level in one of the cultures (C5) of SOT16 protein that catalyzes dsBGLS into BGLS, explains the large accumulation of dsBGLS (Fig. S1). Here, the high level of dsBGLS despite undetectable levels of GGP1 suggests that a yeast γ-glutamyl transpeptidase can substitute for the Arabidopsis GGP1.Conversely, the one-copy genome integration in BGLS$_g$ resulted in less variation in protein level for each enzyme except CYP83B1

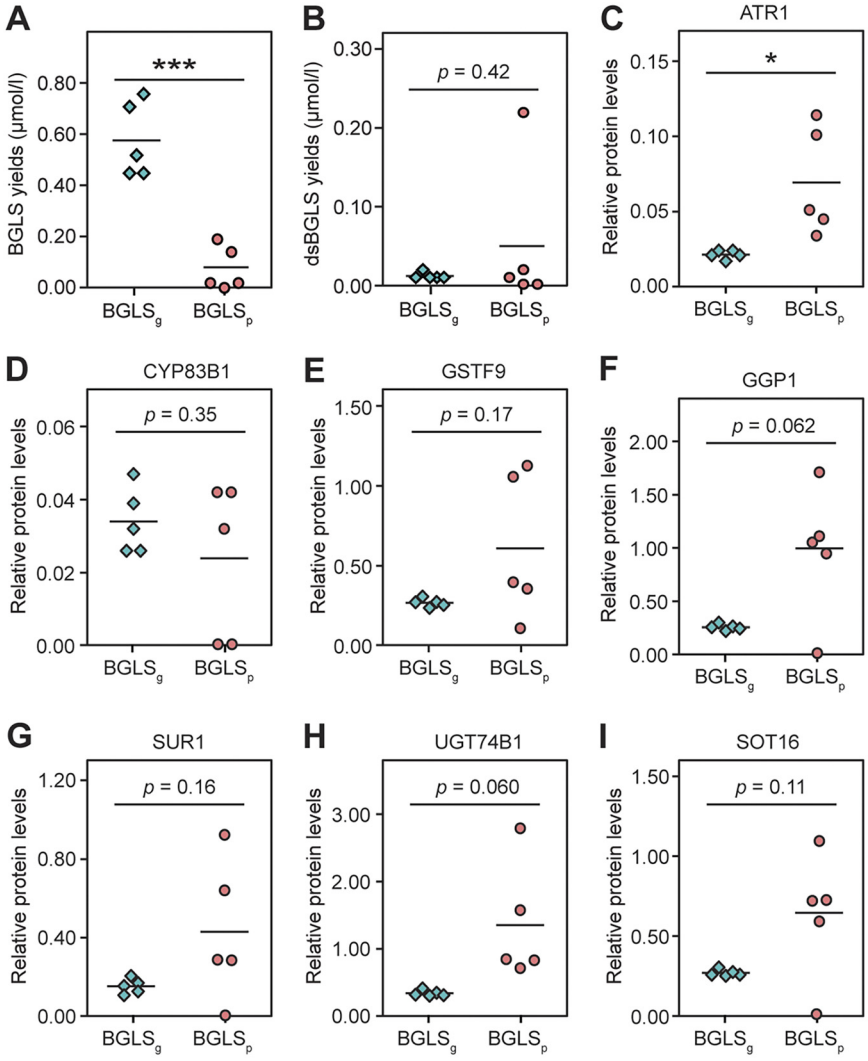

FIG 3 Analysis of BGLS production and protein levels of individual BGLS biosynthetic enzymes in the $BGLS_g$ and $BGLS_p$ strains. (A and B) The yields of BGLS and dsBGLS produced by the $BGLS_g$ strain and the $BGLS_p$ strain. The $BGLS_g$ strain represents the $BGLS_g$ strain transformed with the two empty plasmids pESC-URA-USER and pESC-HIS. $BGLS_p$ strain represents the constructed strain shown in Fig. 2B. Data represent the average and *P* value (two-sided Student's *t* test) of five biological replicates. The *P* value used to compare the BGLS yields produced by the $BGLS_g$ strain and the $BGLS_p$ strain is 0.00039 (indicated with ***). (C to I) Targeted proteomics of protein levels in the individual BGLS biosynthetic enzymes in the $BGLS_g$ and the $BGLS_p$ strains. The spots represent the relative level of a representative peptide for each protein that was normalized to the housekeeping gene *PGI1* (encoding glucose-6-phosphate isomerase). The averages and *P* values (two-sided Student's *t* test) of five biological replicates are shown. The *P* value used to compare the protein levels of ATR1 between the $BGLS_g$ strain and the $BGLS_p$ strain is 0.041 (indicated with *). The engineered strains were grown in SC-Ura-His-Glu media, and samples were taken at 48 h for metabolite analysis and at 24 h for targeted proteomics analysis. Exact values for BGLS and dsBGLS yields and protein levels are listed in Table S1.

(Fig. 3 and Fig. S1). This suggests that genome integration results in more stable gene expression, which is beneficial for engineering the multigene pathway of BGLS.

To investigate whether the highly variable levels of gene expression reflected alteration of presence of full-length genes, the size of the BGLS biosynthetic genes were checked in the $BGLS_g$ and $BGLS_p$ strains by amplification of the whole genes with PCR (Fig. S2). The results show that all the BGLS biosynthetic genes are present and with full length in the $BGLS_g$ strain, which is in accordance with the detected protein levels in Fig. 3 and Fig. S1. For the $BGLS_p$ strain, one of the cultures lacked detectable amounts of *SUR1* and had less abundance of *GGP1*, as also reflected in the absence and reduction in protein levels,

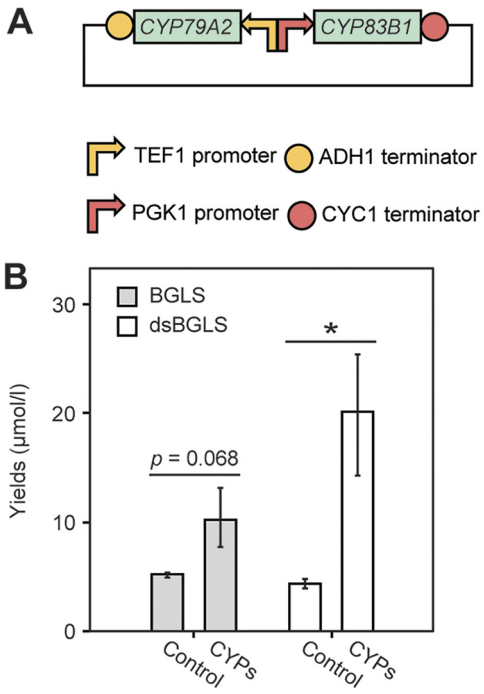

FIG 4 Optimization of BGLS production in $BGLS_g$ strain by enhancing the entry point enzyme activity of CYP79A2 and CYP83B1. (A) Construct for overexpression of *CYP79A2* and *CYP83B1*. The *CYP79A2* and *CYP83B1* genes flanked with the same promoters and terminators shown in Fig. 2 are constructed on the high-copy-number plasmid pESC-URA-USER. (B) Yields of BGLS and dsBGLS produced by the $BGLS_g$ strain overexpressing the entry point genes *CYP79A2* and *CYP83B1*. Control and CYPs represent the $BGLS_g$ strain containing, respectively, the empty plasmid pESC-URA-USER and the construct with *CYP79A2* and *CYP83B1* shown in A. The engineered strains were grown in SC-Ura-Gal medium, and samples were taken at 48 h. Data represent the averages and *P* values (two-sided Student's *t* test) of four biological replicates. The *P* value used to compare the dsBGLS yields produced by the control strain and the CYPs strain is 0.011 (indicated with *). Exact values for BGLS and dsBGLS are listed in Table S1.

respectively. Noticeably, the intensity of DNA fragments representing *CYP83B1* was weak in all cultures of $BGLS_p$ strain. The results suggest that some genes are eliminated by homologous recombination in the $BGLS_p$ strain. This may explain the undetected and variable protein levels.

Optimization of BGLS production in the $BGLS_g$ strain by enhancing the entry point of the BGLS biosynthetic pathway. Using the $BGLS_g$ strain with the biosynthetic genes stably integrated into the genome as platform, we used a 2μ-based plasmid to overexpress the entry point genes *CYP79A2* and *CYP83B1* for increasing BGLS production (Fig. 4A). Interestingly, dsBGLS could only be detected in the culture, and not in yeast cells, whereas the end product BGLS accumulated in small amounts in the cells (Fig. S3). This suggests that dsBGLS is either exported or moved along the pathway, whereas the end product BGLS accumulates to a certain level before it is being exported. Accordingly, we only analyzed the levels of BGLS and dsBGLS in culture in this study. When *CYP79A2* and *CYP83B1* driven by the constitutive promoters TEF1 and PGK1 were introduced into the high-copy-number plasmid pESC-URA-USER and grown in galactose-containing medium minus uracil, we found that the $BGLS_g$ strain produced 10.37 μmol/L of BGLS, which is a 2-fold higher level than a control strain with the empty plasmid pESC-URA-USER (Fig. 4). Under these growth conditions, a similar level of dsBGLS and BGLS was detected in the control strain, whereas a 4.8-fold increase in the production of dsBGLS (20.67 μmol/L) was observed in the strain overexpressing *CYP79A2* and *CYP83B1* (Fig. 4). Overall, this result shows that enhanced expression level of the entry point enzymes CYP79A2 and CYP83B1 increases the BGLS production level and thus the flux through the BGLS pathway. However, the large accumulation of

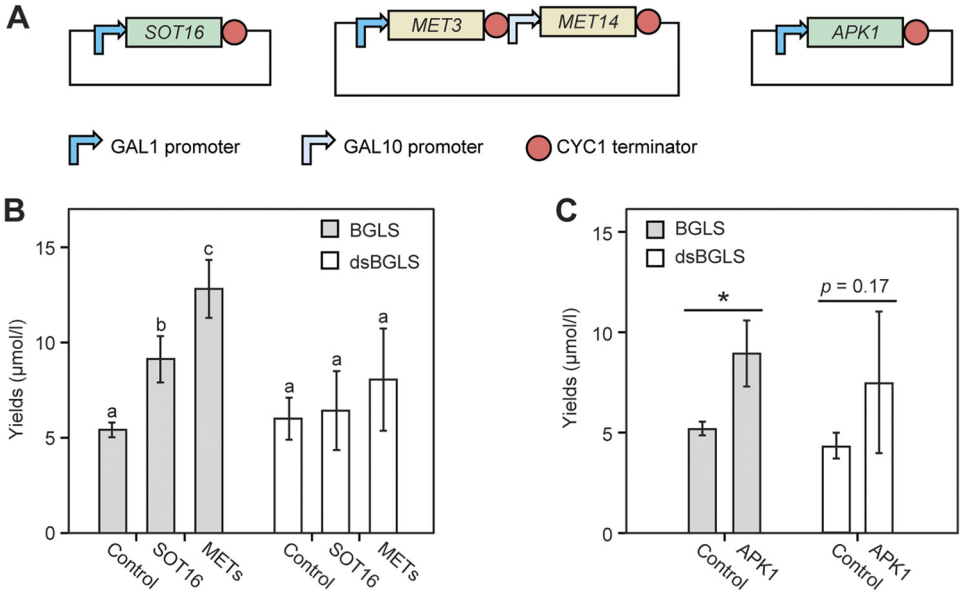

FIG 5 Alleviation of metabolic bottleneck in the conversion of dsBGLS to BGLS in the $BGLS_g$ strain. (A) Constructs for overexpression of *SOT16*, *MET3*, and *MET14*, and expression of *APK1*. The *SOT16* gene and *APK1* gene encoding *A. thaliana* APS kinase were constructed on plasmids pESC-HIS and pESC-URA-USER, respectively. The *MET3* and *MET14* were constructed on plasmid pESC-HIS. Expression of all the genes was controlled by the terminator CYC1 and the galactose-inducible promoter GAL1, except expression of *MET14* under the control of the galactose-inducible promoter GAL10. (B) Yields of BGLS and dsBGLS from the $BGLS_g$ strain overexpressing *SOT16* or *MET3* and *MET14*. The abbreviations Control, SOT16, and METs represent the $BGLS_g$ strain containing, respectively, the empty plasmid pESC-HIS, the construct with *SOT16*, and the construct with *MET3* and *MET14*. The engineered strains were grown in SC-His-Gal medium, and samples were taken at 48 h. Statistical significance is shown based on Pairwise Tukey's HSD test ($P < 0.05$), and data labeled with different letters are significantly different. Exact values for BGLS and dsBGLS yields and *P* values are listed in Table S1. (C) Yields of BGLS and dsBGLS from the $BGLS_g$ strain overexpressing *APK1*. Control and APK1 represent the $BGLS_g$ strain containing, respectively, the empty plasmid pESC-URA-USER and the construct with *APK1*. The engineered strains were grown in SC-Ura-Gal medium, and samples were taken at 48 h. Data represent the average and *P* values (two-sided Student's *t* test) of four biological replicates. The *P* value used to compare the BGLS yields produced by the control strain and the APK1 strain is 0.015 (indicated with *). Exact values for dsBGLS and BGLS yields are listed in Table S1.

dsBGLS indicates that the last step of the BGLS pathway is a metabolic bottleneck when the $BGLS_g$ strain is growing under these experimental conditions in this medium.

Alleviation of the metabolic bottleneck at the sulfotransferase step in the BGLS production in $BGLS_g$ strain. Towards alleviating the metabolic bottleneck at the last step in the BGLS pathway catalyzed by SOT16, we first enhanced the expression of the *SOT16* gene by increasing copy number and using the strong galactose-inducible promoter GAL1 to regulate the expression (Fig. 5A). When we monitored the production level of BGLS and dsBGLS at 48 h after induction, we observed that the BGLS yield significantly increased by 1.7-fold to 9.12 μmol/L (Fig. 5B and Table S1). However, no significant difference in dsBGLS accumulation was observed by overexpression of *SOT16* (Fig. 5B), showing that although the elevated expression level of *SOT16* improved BGLS production, the metabolic bottleneck remained. The results suggest that an increased availability of the SOT16 enzyme required an increased availability of the sulfotransferase 3′-phosphoadenosine-5′-phosphosulfate (PAPS) cofactor that could be a limiting factor.

In *S. cerevisiae*, PAPS is an intermediate in the sulfate assimilatory pathway (14). To increase PAPS availability, we overexpressed two genes, *MET3* and *MET14*, encoding, respectively, ATP sulfurylase and adenylylsulfate kinase, i.e., the first two enzymes converting sulfate to PAPS in the sulfate assimilation pathway of yeast (14). The strong galactose-inducible promoters GAL1 and GAL10 were used to control the expression of *MET3* and *MET14*, respectively (Fig. 5A). We found that the BGLS level significantly increased by 2.4-fold, resulting in a yield at 12.81 μmol/L, whereas the level of dsBGLS was similar to the control level, thus showing that the metabolic bottleneck was not

alleviated (Fig. 5B). As an alternative strategy to boost PAPS generation, we introduced an Arabidopsis adenosine 5-phosphosulfate kinase, the APS kinase APK1, which is an ortholog of MET14. Expression of *APK1* driven by the galactose-inducible promoter GAL1 on the high-copy-number plasmid pESC-URA-USER resulted in an increased production of BGLS by 1.7-fold, whereas the level of dsBGLS did not change (Fig. 5C). The result shows that introduction of APK1 increases the BGLS production even though the sulfotransferase step is still a metabolic bottleneck.

Interestingly, combining overexpression of *CYP79A2*/*CYP83B1* together with overexpression of *SOT16* or *MET3*/*MET14*, which required that we use medium with SC-Ura-His-Gal, showed that the increased levels of BGLS from individually overexpressing *CYP79A2*/*CYP83B1*, *SOT16*, or *MET3*/*MET14* were not observed (Fig. S4). The observed difference in effect of adding *SOT16* and *MET3*/*MET14* in *CYP79A2*/*CYP83B1* overexpressed lines may reflect the introduction of additional plasmids as well as the different medium, thereby demonstrating the challenges of engineering multigene pathways in microbes.

DISCUSSION

In this study, we used *S. cerevisiae* as a microbial cell factory to engineer BGLS production by either genome engineering (strain $BGLS_g$) or plasmid-based engineering ($BGLS_p$) of the biosynthetic genes. The $BGLS_g$ strain produced 8.4-fold higher levels of BGLS than $BGLS_p$ (Fig. 3), despite a generally lower protein level of the individual biosynthetic enzymes in $BGLS_g$ compared to $BGLS_p$. The various protein level of the individual enzymes in $BGLS_p$ strain—in some samples even undetectable levels—may explain the overall low BGLS yield in this strain. Generally, the copy number of 2μ-based plasmids varies from cell to cell (15), causing many different combinations when more than one kind of plasmid is present in a cell. Thus, cells with a balanced production of all the enzymes might be rare, especially for pathways with multiple genes. For example, only 6% of the yeast cells possess identical levels of the three fluorescent proteins, YFP, RFP, and CFP, coexpressed from three different 2μ-based plasmids (16). However, integration of the three genes into the yeast genome results in 95% of the cells with identical protein levels. Their findings are consistent with our protein level analysis for enzymes in the BGLS pathway engineered by genome integration and 2μ-based plasmid expression (Fig. 3 and Fig. S1). The problem of multiple, different 2μ-based plasmids may be alleviated through constructing the entire pathway into a single plasmid.

Another possible explanation is the low genetic stability of plasmids in yeast. For example, two fluorescent proteins, CFP and RFP, expressed from 2μ-based plasmids showed that only 54% of the cells produced both proteins at midexponential phase (17). Noticeably, when single genes were expressed, the protein level produced from 2μ-based plasmids exceeded that from genome integration (17). Another study reported that only around 28% of the yeast cells transformed with two 2μ-based plasmids containing three genes for synthesizing artemisinic acid maintained both of the plasmids after 120 h growth (18). Furthermore, copy numbers affect plasmid stability, as evidenced by a study expressing cheilanthifoline synthase from *Eschscholzia californica* from a low- and high-copy-number plasmid, respectively: 90% of the low-copy-number plasmid was retained while only 10.7% of the high-copy-number plasmid was maintained (10). Although strains with high-copy plasmids have stability problems, utilization of these plasmids is still a common method to achieve high level of gene expression for metabolic engineering (19).

In *S. cerevisiae*, plasmids containing direct repeat are known to be unstable due to intraplasmid direct repeat recombination leading to loss of DNA between the repeats (20). Accordingly, plasmids supporting a multigene pathway such as BGLS biosynthesis may be prone to gene loss if some of the genes are equipped with the same promoter and terminator. In our case, the plasmids contain four genes, and the central two genes are flanked by a direct repeat constituted by a pair of identical bidirectional promoters (Fig. 2B). For example, the undetected GGP1 and SOT16 in culture 5 may be caused by intraplasmid direct repeat recombination between the two identical bidirectional promoters TEF1 and

PGK1 (Fig. 2B and Fig. S1). Similarly, the lack of detection of CYP83B1 and SUR1 in culture 3 may be the result of interplasmid direct repeat recombination between the terminator ADH1 and the promoter TEF1 or between the promoter PGK1 and the terminator CYC1 (Fig. 2B and Fig. S1). Such events may be quite frequent as high transcription levels are accompanied by increased direct repeat recombination rates (21). Since each direct repeat exists in a pool of multicopy plasmids, direct repeat recombination may appear more frequently in plasmids than for corresponding direct repeats positioned in single copies at the various integration sites in the genome. To this end, we note that selection will favor recombinants that have lost the ability to produce potentially toxic intermediates and end products. Hence, the plasmids that result from direct repeat recombination may take over the population not only as they are genetically more stable and easier to replicate due to the smaller size, but also since they cause less metabolic stress on the cells. For comparison, all the enzymes in BGLS biosynthesis were present in stable levels in the $BGLS_g$ strain. As a strategy to increase copy number of stably integrated genes, the genome engineering platform used in the current study has been further advanced in a gene amplification system, CASCADE, which allows amplicons containing one or more genes (i.e., entire multigene pathways) to be efficiently introduced into the genome with defined numbers of amplicons ranging from one to nine integrated copies (17).

Other work has also been reported to compare genome integration and multicopy plasmids for engineering multigene pathways in *S. cerevisiae*. For example, a naringenin biosynthetic pathway containing six genes was constructed into three high-copy plasmids, and the amount of target product exhibited large variability and low reproducibility among transformant colonies (22). The strain harboring a genomically integrated pathway showed superior reproducibility, although a lower level of naringenin was produced.

Optimization of BGLS production with the $BGLS_g$ strain through overexpression of the entry point enzymes (CYP79A2 and CYP83B1) enhanced the production level of BGLS and thus increased the flux through the pathway (Fig. 4). This is consistent with a previous study in *E. coli* in which enhanced enzyme activity of CYP79A2 increases BGLS production (3). The increased flux through the pathway revealed a metabolic bottleneck at the last step catalyzed by the sulfotransferase SOT16 resulting in dsBGLS accumulation (Fig. 4). This result is different from the tiny amount of dsBGLS yield by the $BGLS_g$ strain grown in glucose-containing synthetic media (Fig. 3B). This may reflect the presence of either one or two plasmids transformed into $BGLS_g$ strain and thus different drop-out media. This emphasizes the importance of the number of plasmids introduced and the experimental growth conditions in pathway engineering.

Possible explanations of the bottleneck could be inefficient enzyme level of sulfotransferase SOT16 or inadequate supply of cofactor PAPS. The bottleneck was partly alleviated by overexpression of *SOT16* as well as the endogenous *MET3* and *MET14* and the Arabidopsis APK1 (Fig. 5). Accumulation of dsBGLS in the media was also observed when engineering BGLS production in *E. coli* and *N. benthamiana* (3 and 23). In *E. coli*, overexpression of three native PAPS-generating genes encoding an adenylyl-sulfate kinase and two subunits of a sulfate adenylyltransferase did not increase, but reduced BGLS production even when supplemented with different sulfur sources in media (3). In tobacco, coexpression of APS kinase 2, APK2, reduced the majority of dsBGLS accumulation and increased the BGLS level by 16-fold, presumably through increased regeneration of PAPS (23). Furthermore, screening biodiversity by testing other sulfotransferases did not improve BGLS production level in either tobacco or *E. coli* (3 and 23). Despite the partial alleviation of metabolic bottleneck, the problem of the dsBGLS accumulation remained. This could be due to deficient PAPS level, since the intracellular levels of PAPS have been reported to be fine-tuned in the microorganisms due to toxicity (24). Moreover, when PAPS has donated a sulfate group to dsBGLS, adenosine-3′,5′-bisphosphate (PAP) accumulates, and accumulation of PAP has been reported to inhibit sulfotransferases and be toxic to yeast (25). Additionally, export of dsBGLS from

TABLE 1 Yeast strains used in this study

Strain abbreviations	Description	Reference
CEN.PK 113-11C	The original strain for gene expression (*MATa MAL2-8C SUC2 his3Δ ura3-52*).	Peter Köttera[a]
$BGLS_g$	Genome-engineered strain: with *CYP79A2*, *CYP83B1*, *GSTF9*, *SUR1*, *GGP1*, *UGT74B1*, *SOT16*, and *ATR1* integrated in the XII chromosome, shown in Fig. 2A.	This study
$BGLS_p$	Plasmid-engineered strain: harboring multiplasmids pESC-URA-USER with *CYP79A2*, *CYP83B1*, *GSTF9*, and *ATR1*, and pESC-HIS with *SUR1*, *GGP1*, *UGT74B1* and *SOT16*, shown in Fig. 2B.	This study
CYPs	The $BGLS_g$ strain harboring the construct pESC-URA-USER with *CYP79A2* and *CYP83B1*, shown in Fig. 4B.	This study
SOT16	The $BGLS_g$ strain harboring the construct pESC-HIS with *SOT16*, shown in Fig. 5B.	This study
METs	The $BGLS_g$ strain harboring the construct pESC-HIS with *MET3* and *MET14*, shown in Fig. 5B.	This study
APK1	The $BGLS_g$ strain harboring the construct pESC-URA-USER with the truncated *APK1* without chloroplast localization sequences, shown in Fig. 5C.	This study

[a]Institut für Mikrobiologie, der Johan Wolfgang Goethe-Universität, Frankfurt am Main, Germany.

the cells could be a means of detoxification leading to premature abortion of the BGLS pathway.

Interestingly, comparison of the levels of BGLS production in *E. coli* and *S. cerevisiae* shows that *E. coli* produced the highest level with 20.3 μmol/L (equivalent to 8.3 mg/L) (3), whereas the highest level of BGLS produced by yeast is 12.81 μmol/L (=5.2 mg/L) (this study). However, the engineered yeast strain in this study may have better BGLS production potential since it is grown in selection-free medium, while the engineered *E. coli* strain produces BGLS dependent on two antibiotics (3). For both microbial host organisms, additional strain optimization is required to reach commercial levels.

In summary, this study reports the first comparison of the production in yeast of a plant natural product, BGLS, by engineering a multigene pathway from high-copy plasmids or from stable one-copy integration in genome. The 8.4-fold higher production of BGLS in the $BGLS_g$ strain combined with consistent protein levels demonstrates that genome integration is the preferred strategy for engineering of multigene biosynthetic pathways.

MATERIALS AND METHODS

Strain background and media. All the BGLS-producing strains were constructed based on the *S. cerevisiae* strain CEN.PK 113-11C (MATa MAL2-8C SUC2 his3Δ ura3-52), listed in Table 1.

The synthetic complete (SC) drop-out media were used to grow the engineered yeast strains. The media include 20 g/L carbon source glucose (Glu) or galactose (Gal), 6.7 g/L Yeast Nitrogen Base Without Amino Acids (Sigma-Aldrich), various concentrations of Yeast Synthetic Drop-out Medium Supplements (Sigma-Aldrich), and 20 g/L agar for transformation or without agar for culturing. The specific Yeast Synthetic Drop-out Medium Supplements for each medium are described in Table S2.

Generation of genome-engineered $BGLS_g$ strain and plasmid-based $BGLS_p$ strain. Four expression cassettes containing the genes of the BGLS biosynthetic pathway (*CYP79A2*, *CYP83B1*, *GSTF9*, *SUR1*, *GGP1*, *UGT74B1*, and *SOT16*; see Fig. 1) and the cytochrome P450 electron-donating support gene *ATR1* were inserted into the selected integration sites of the genome following a previously described approach (7). All the genes used in this study are listed in Table 2. A bidirectional constitutive *TEF1/PGK1*-promoter and terminators ADH1 and CYC1, respectively, were cloned into the plasmids to control the gene expression. The four integration cassettes were as follows: pXII-1 containing *CYP79A2* and *CYP83B1*, pXII-2 containing *GSTF9* and *ATR1*, pXII-3 containing *SUR1* and *GGP1*, and pXII-4 containing *SOT16* and *UGT74B1*. For schematic illustration of the genome integration strategy, see Fig. 2A. To inte-

TABLE 2 The genes used to engineer and optimize BGLS production

Gene	Enzyme	Species	Locus
CYP79A2	Cytochrome P450	*Arabidopsis thaliana*	AT5G05260
CYP83B1	Cytochrome P450	*Arabidopsis thaliana*	AT4G31500
GSTF9	Glutathione S-transferase	*Arabidopsis thaliana*	AT2G30860
GGP1	γ-glutamyl peptidase	*Arabidopsis thaliana*	AT4G30530
SUR1	C-S lyase	*Arabidopsis thaliana*	AT2G20610
UGT74B1	Glycosyltransferase	*Arabidopsis thaliana*	AT1G24100
SOT16	Sulfotransferase	*Arabidopsis thaliana*	AT1G74100
ATR1	P450 reductase	*Arabidopsis thaliana*	AT4G24520
APK1	APS kinase	*Arabidopsis thaliana*	AT2G14750
MET3	ATP sulfurylase	*Saccharomyces cerevisiae*	YJR010W
MET14	Adenylylsulfate kinase	*Saccharomyces cerevisiae*	YKL001C

grate the eight genes on the genome, the constructs pXII-1, pXII-2, pXII-3, and pXII-4 were digested by XbaI, and the four linear DNA fragments were individually and iteratively transformed into the yeast strain CEN.PK 113-11C, generating a strain named $BGLS_g$. The *URA3* marker gene was eliminated by direct repeat recombination and 5-fluoroorotic acid (5-FOA) selection to prepare the strain for the next transformation (26). All the primers used to generate the constructs are listed in Table 3.

Based on the four integration constructs, the eight genes flanked with the same promoters and terminators were cloned into the 2μ-based high-copy plasmids pESC-URA-USER and pESC-HIS (number 217451, Agilent Technologies), resulting in pESC-URA-USER harboring *CYP79A2*, *CYP83B1*, *GSTF9*, and *ATR1* and pESC-HIS harboring *SUR1*, *GGP1*, *SOT16*, and *UGT74B1*. For schematic illustration of the constructs, see Fig. 2B. The plasmid pESC-URA-USER was previously generated by adding USER cassette to the plasmid pESC-URA (number 217454, Agilent Technologies) for USER cloning (27). The two plasmids with the eight genes for BGLS production were simultaneously transformed into CEN.PK 113-11C, generating a strain named $BGLS_p$. All yeast strains were generated using the previously described transformation method (28). Integration and transformation were confirmed by colony PCR.

Strategies for improvement of BGLS production in $BGLS_g$ strain. To optimize the BGLS production level, the expression cassette containing the genes *CYP79A2* and *CYP83B1* were amplified by PCR using the integration construct pXII-1 as the template and inserted into the plasmid pESC-URA-USER. The gene *SOT16* was constructed into the plasmid pESC-HIS, flanked with galactose-inducible promoter GAL1 and terminator CYC1. The yeast native genes *MET3* and *MET14* were cloned into the plasmid pESC-HIS under the control of galactose-inducible promoters GAL1 and GAL10, respectively. The chloroplast localization sequences (29) were removed from the *APK1* gene of *A. thaliana*, and the truncated gene was amplified from cDNA. The gene was constructed into the plasmid pESC-URA-USER.

Growth conditions of yeast strains. For generation of yeast strains containing the BGLS biosynthetic genes in the genome, the transformants were grown with SC-Ura-Glu agar medium. Subsequently, the strains were grown on synthetic complete media containing 30 mg/L uracil and 740 mg/L 5-FOA to excise the *URA3* maker gene. For transformation of pESC-URA-USER-based constructs and pESC-HIS-based constructs, the transformants were grown with SC-Ura-Glu and SC-His-Glu agar media, respectively. Similarly, SC-Ura-His-Glu agar medium was used to select the yeast strain transformed with both pESC-URA-USER-based and pESC-HIS-based constructs.

All the strains were cultured in 100 mL baffled flasks containing 30 mL media at 30°C, 150 rpm. Glucose was used as the carbon source to grow all the precultures and the first BGLS-producing cultures for comparison of genome integration and plasmid-based introduction of the BGLS biosynthetic genes. Galactose was used as the carbon source to grow the BGLS-producing culture for optimization of BGLS production. Single colonies were inoculated in media for around 24 h as preculture. OD_{600} values of each preculture were measured, and volumes of each preculture were calculated for the original OD_{600} to 0.6 in 30 mL fresh media. After the calculated volume of preculture was precipitated at 5,000 $\times$ *g* for 5 min, the cells were suspended in fresh media as expression culture. Cultures from 24 h growth were used for analysis of targeted proteomics, and cultures from 48 h growth were used for metabolic analysis.

Metabolite extraction and LC-MS analysis. The expression culture was precipitated at 17,000 $\times$ *g* for 5 min to collect the supernatant. The supernatant was diluted by 10-fold with a buffer containing 10 μg/mL 13C- and 15N-labeled amino acids (Algal amino acids 13C and 15N, Isotec, Miamisburg, USA) and 2 μmol/L sinigrin (PhytoLab, Vestenbergsgreuth, Germany). Subsequently, the diluted samples were filtered (Durapore 0.22-μm PVDF filters, Merck Millipore, Tullagreen, Ireland) and used directly for LC-MS analysis. Sinigrin and 13C, 15N-Phe were used as internal standard to quantify BGLS and dsBGLS, respectively.

BGLS, dsBGLS, and the level of amino acids were monitored by LC-MS analysis using a modified version of a previously described method (30). Briefly, an Advance UHPLC system (Bruker, Bremen, Germany) was used for chromatography. A Zorbax Eclipse XDB-C18 column (100 $\times$ 3.0 mm, 1.8 μm, Agilent Technologies, Germany) was used for separation. The mobile phases A and B were formic acid (0.05% [vol/vol]) in water and acetonitrile, respectively. The elution profile was as follows: 0 to 1.2 min 3% B; 1.2 to 4.3 min 3 to 65% B; 4.3 to 4.4 min 65 to 100% B; 4.4 to 4.9 min 100% B, 4.9 to 5.0 min 100 to 3% B; and 5.0 to 6.0 min 3% B. Mobile phase flow rate was 500 μL/min with the column temperature maintained at 40°C. The liquid chromatography was coupled to an EVOQ Elite TripleQuad mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ionization source (ESI). Infusion experiments with pure standards were used to optimize instrument parameters. Ionspray voltage was set to 3,000 V or $-$4,000 V in positive and negative ionization mode, respectively. Cone temperature was maintained at 350°C and cone gas (nitrogen) flow to 20 lb/in^2. Heated probe temperature was maintained at 400°C, and probe gas flow to 50 lb/in^2. Nebulizing gas (nitrogen) was set to 60 lb/in^2 and collision gas (argon) to 1.6 mTorr.

Multiple reaction monitoring (MRM) was used to monitor analyte the parent ion to product ion transitions: MRM transitions for 13C-, 15N-labeled phenylalanine was used as previously reported (31). MRM transitions for BGLS, dsBGLS, and sinigrin were used as previously reported (32). Q1 and Q3 quadrupoles were maintained at unit resolution. Data acquisition and processing were achieved by Bruker MS Workstation software (Version 8.2.1, Bruker, Bremen, Germany). For analytes, multiple transitions were monitored and the transition used for quantification was marked as quantifier (Qt). The information of transitions and collision energies is seen in Table S3. Dilution series of the respective analytes were used to calculate response factors to the respective internal standards. The selection of internal standards is based on matching ionization mode with the analyte of interest (i.e., BGLS in negative ionization mode; dsBGLS and amino acids in positive ionization mode). Due to the matrix effect from the cultures on the quantification of the analytes, the correction factors were calculated into the response factors in Table S4.

TABLE 3 The primers used to generate the constructs and verify genes in this study

Primers	Sequences
BGLS-1	TCCCAGATTTGGCTTTGATT
BGLS-2	GCTCATTAGAAAGAAAGCATAGCA
BGLS-3	GCCACGTGCTTTATGAGGGT
BGLS-4	AGATCACCGCGAGGCGAC
BGLS-5	TTGCGTCTGCGATAGTTTC
BGLS-6	TCAATTGAGATGAGCTTAATCATGT
BGLS-7	CGGGCAATCAGAATCTGTAAC
BGLS-8	TCATATCTCGCTTTGATTTCTTCGG
BGLS-9	CATAGGCCGCCAAGGCAA
BGLS-10	GTCCCAGGTCCCACCATAA
BGLS-11	GGAAGCTAAGGATGGTTG
BGLS-12	CTTACCCTGGCTATGATCTG
BGLS-13	AGCCGCTGTAGCTACTTAAG
BGLS-14	TTATCATTTCTATTATTATCCTGCTCAGT
BGLS-15	GTAGATAATTACTTCCTTGATGATCTG
BGLS-16	CGAACCAAACGAAACAAATGCT
BGLS-17	CTTGATTGGAGACTTGACCAAACCTCTGGCGAAGAATTGTTAATACGGAATGCGTGCGAT
BGLS-18	TCTCAGGTATAGCATGAGGTCGCTCCTTCGAGCGTCCCAAAACCTTCT
BGLS-19	TGAGAAGGTTTTGGGACGCTCGAAGGAGCGACCTCATGCTATAC
BGLS-20	TTTTCGGTTAGAGCGGATCTTAGCTAGCCGCGGTACCAAGCTCTACCAGACATCTCTGAG
BGLS-21	TGATTGGAGACTTGACCAAACCTCTGGCGAAGAATTGTTAATATGCGTGCGATTCACTAC
BGLS-22	AGAGCGGATCTTAGCTAGCCGCGGTACCAAGCTTACTCGACTACTTCCCTAAACTCTCTATAAACT
BGLS-23	ATGCTCGCGTTTATTATAGG
BGLS-24	TCGGTGTTGAGTTCTTTCAT
BGLS-25	TGCAGCCGCTACTAAAC
BGLS-26	GTGAATTTGATGGAGAAAGG
BGLS-27	GCAAGCACTTTTGGGCACT
BGLS-28	CCTGCTTATCTCGCTCTAC
BGLS-29	TCCCATGACTGATGCGAAC
BGLS-30	CAAGTACAACATCGTCGG
BGLS-31	GGCCACATTTGATTCCATT
BGLS-32	GTAAGAAAGGTGCAAAGCGT
BGLS-33	ATGAGCGAAGAACAACCAC
BGLS-34	GGAAGCAAGGGTAGACGG
BGLS-35	CGCTGACGACTCTAAAGG
BGLS-36	GGAACTCTGCCTTTGAGGAA
BGLS-37	ACGAGACCATGAGGGCCAAT
BGLS-38	AATCAAAGACAACCCAAAAC
BGLS-39	GCTACAGACGAGGAAATCAG
BGLS-40	GGTGTTGTGATTGAGGAG
BGLS-41	CTTTAATCACCCACAAGAAGTTC
BGLS-42	TCGCTCTTAATTAACTTCGAGCGTCCCAAAACCT
BGLS-43	TCGAAGGCGGCCGCAGAGGCGGTTTGCGTATTGG
BGLS-44	TCGAAGTTAATTAAGAGCGACCTCATGCTAT
BGLS-45	GCCTCTGCGGCCGCCTTCGAGCGTCCCAAAAC
BGLS-46	CCCCGCTTAATTAAGCGTTGGCCGATTCATT
BGLS-47	TCGAAGGCGGCCGCCGCTCACTGACTCGCT
BGLS-48	CAACGCTTAATTAAGAGCGACCTCATGCTATACC
BGLS-49	TGAGCGGCGGCCGCCTTCGAGCGTCCCAAAACCT
BGLS-50	CCGTATTACCGCCTTTGAG
BGLS-51	ATAGGGCCCGGGATGGAATCAAAGACAACCCAAAAC
BGLS-52	AGCTTACTCGAGTCAGTTATCATGTTGAAGCAAGCCAGT
BGLS-53	GCCGTTUATGAGCACGAACATCAAATGGCACGA
BGLS-54	GCTGTTUTTAGGCCTGTAGATAGCCTTTATTGT
BGLS-55	GCCGTTUATGCCTGCTCCTCACGGTGGT
BGLS-56	GCTGTTUTTACAAATGCTTACGGATGATTTTTTCACTGAT
BGLS-57	ACCGTGTCGTAGCTTTCCAA
BGLS-58	AGATTTCATTGGTTCTGGTAGT
BGLS-59	AACCCCGGATCCATGCCTGCTCCTCACGGTG
BGLS-60	CAAGCTTACTCGAGTTACAAATGCTTACGGATGATTTTTT
BGLS-61	ACGGAATGCGTGCGATTCACT
BGLS-62	ATGCTCGCGTTTATTATAGGTTTGCTTCT

(Continued on next page)

TABLE 3 (Continued)

Primers	Sequences
BGLS-63	CGTAATACTATGGATCTCTTATTGATT
BGLS-64	CGTTGGTGCAAGAACGAGG
BGLS-65	TGCGTTAAGCTGGGAATGAATA
BGLS-66	ATGGTGCTAAAGGTGTACGGA
BGLS-67	ATGACTTCTGCTTTGTATGCT
BGLS-68	TACCAGACATCTCTGAGGTATCTT
BGLS-69	CATTTCGAGATTATTATCACTCAGTT
BGLS-70	ATGAGCGAAGAACAACCAC
BGLS-71	ATGGTGGAGCAAAAGAGATA
BGLS-72	GGAACTCTGCCTTTGAGGAA
BGLS-73	TTATCATGTTGAAGCAAGCCA
BGLS-74	AATCAAAGACAACCCAAAAC
BGLS-75	ATGGCGGAAACAACTCCCAA
BGLS-76	CTACTTCCCTAAACTCTCTATAAACT

Targeted proteomics. For relative quantification of targeted proteins, a set of stable isotopically labeled proteotypic peptides were ordered as internal reference (JBL, SpikeTides). Peptides for the BGLS biosynthetic enzymes were designed and kindly given by Meike Burrow and Daniel Vik, except the peptides for ATR1, GSTF9, and the yeast housekeeping enzyme PGI1 (locus: YBR196C), which were designed by Annette Petersen. Skyline 4.2 was used to predict tryptic digestion *in silico* for the possible peptides of each enzyme (33). At least two nonneighboring promising peptides were selected for each target protein analysis according to these rules: *m/z* below 1,250, cleavage sites without neighboring lysine or arginine, and as little methionine and cysteine as possible. The specific peptides for each enzyme are listed in Table 4.

A combined optimized method was used to extract proteins from yeast cells for quantitative proteomics (34, 35). Specifically, the yeast cells from 10 mL culture were resuspended in 300 μL lysis buffer (0.1 M NaOH, 0.05 M EDTA, 2% SDS, 2% 2-mercaptoethanol) followed by incubation at 90°C for 10 min. After 5 μL of 4 M acetic acid was added, the solution was vortexed for 30 s and incubated at 90°C for 10 min. For phase separation, 600 μL methanol, 150 μL chloroform, and 450 Milli-Q water were successively added, and the samples were centrifuged at 4°C, 21,000 $\times$ *g* for 1 min. Subsequently, the aqueous phase was removed, and the samples were centrifuged again at 4°C, 21,000 $\times$ *g* for 1 min after another 450 μL methanol was added to dislodge the protein pellets. The proteins were dried using Speed-Vac under the conditions of 1,000 rpm, 30 to 60 min, and maximum 35°C.

The remaining procedure was very similar to the previously published method (3) except the amount of protein for tryptic digestion. Briefly, the protein pellets were resuspended with 100 μL of 100 mM ammonium bicarbonate buffer containing 10% (vol/vol) methanol. According to the protein concentration measured with Pierce BCA Protein assay kit (ThermoFisher, number 23225), 100 μL of 1 mg/mL protein samples (100 μg in total) were prepared for the tryptic digestion. Prior to the digestion, the proteins were incubated for 30 min at room temperature with the addition of 2 μL 10-mM dithiothreitol (DTT) and subsequently incubated in the dark for 20 min at room temperature with the addition of 25 μL 50-mM iodoacetamide. Another 100 μL of 100-mM ammonium bicarbonate buffer containing 10% (vol/vol) methanol was added before incubation at 37°C overnight with 2 μg trypsin/Lys-C mix (Promega, number V5073). The reaction was stopped by acidification with trifluoroacetic acid (TFA).

The peptide digests diluted up to 1.5 mL with buffer A (2% [vol/vol] acetonitrile, 0.1% [vol/vol] formic acid) were purified with Sep-PakC-18 columns (Waters, Sep-Pak Vac 1 cc 100 mg, number WAT023590). The peptide digests were washed three times with 1 mL buffer A and eluted twice with 0.5 mL buffer B (65% [vol/vol] acetonitrile, 0.1% [vol/vol] formic acid). The peptides were dried in Speed-Vac under the condition of 1,000 rpm, 3 to 6 h, and maximum 35°C. For LC-MS analysis, the dry peptides were resuspended in 25 μL buffer C (2% [vol/vol] acetonitrile, 0.5% [vol/vol] formic acid, and 0.1% [vol/vol] trifluoroacetic acid [TFA]) containing 20 nM isotopically labeled peptide standards (JPT, SpikeTides) and filtered through 0.22-μm centrifugal filters (Merck, number UFC30GV00).

Liquid chromatography was performed on a 1290 Infinity II UHPLC system (Agilent Technologies). Gradient conditions were adopted, with changes, from Percy et al. and Batth et al. (36, 37) and modified from previous studies on the pathway by Petersen et al. (3). Briefly, MilliQ-grade water with 0.05% formic acid and acetonitrile with 0.05% formic acid were used as solvents A and B, respectively. Flow rate was 400 μL/min with the following gradient conditions: 0.0 to 0.5 min 5% B; 0.5 to 22.0 min 5 to 32% B; 22.0 to 24.0 min 32 to 42% B; 24 to 25 min 42 to 90% B; 25.0 to 28.0 min 90% B; 28.0 to 31.0 min 90 to 5% B; 31.0 to 35.0 min 5% B. Column oven temperature was maintained at 55°C. Peptide separation was achieved on an Aeris PEPTIDE XB-C18 column (150 $\times$ 2.1 mm, 2.6 μm, 100 Å, Phenomenex). The injection volume was 10 μL. The column temperature was maintained at 55°C. The liquid chromatography was coupled to an Ultivo Triplequadrupole mass spectrometer (Agilent Technologies) equipped with a Jetstream electrospray ion source (ESI) operated in positive ion mode. Instrument parameters were optimized by running different mixes of the target peptides. Mixes of the target peptides were also used to identify the three most abundant transitions from analyte precursor ion to fragment ion for multiple reaction monitoring (MRM). Three to five peptides were chosen for detection of the individual proteins. MRM transitions and further details can be found in Table S5. Source settings for heated electrospray

TABLE 4 Proteotypic peptides used as reference for quantification of BGLS biosynthetic enzymes and a yeast housekeeping enzyme in targeted proteomics analysis[a]

Gene	Peptide sequence	Reference
CYP79A2	SWPLIGNLPEILGR	3
	LVIESDLPNLNYVK	3
	LIQGFTWLPVPGK	3
CYP83B1	**GYVSEEDIPNLPYLK**	3
	GQDFELLPFGSGR	3
	LAVISSAELAK	3
GSTF9	LAGVLDVYEAHLSK	3
	GVAFETIPVDLMK	This study
	VYGPHFASPK	This study
	HVSAWWDDISSR	
GGP1	**DAITPGSYFGNEIPDSIAIIK**	3
	VVSGEFPDEK	This study
	YALFLATLDSEFVK	This study
	ILGICFGHQIIAR	
SUR1	FASIVPVLTLAGISK	3
	IGWIALNDPEGVFETTK	3
	EENLVFLPGDALGLK	3
UGT74B1	**GLPSLSYDELPSFVGR**	3
	SINEFIESLGK	This study
	FSNGDFPLPADPNSAPFR	
SOT16	VGDWANYLTPEMAAR	3
	YDDAANPLLK	This study
	YQDFIATLPK	This study
	ALTYAIVNR	
ATR1	VIDLDDYAADDDQYEEK	3
	DEDDDLDLGSGK	This study
	VSIFFGTQTGTAEGFAK	This study
	SVATPYTAVIPEYR	This study
	VHVTSALVYGPTPTGR	
PGI1	**TFTNYDGSK**	This study
	VVDPETTLFLIASK	This study
	TFTTAETITNANTAK	This study
	AEGATGGLVPHK	This study

[a]The peptides in bold are the representative peptides for each protein, which are used for calculation of the relative protein level.

ionization were as follows: spray voltage 3,000 V, positive ion mode; gas temperature 325°C; gas flow 13 L/min; nebulizer 25 lb/in^2; sheath gas temperature 400°C; and sheath gas flow 12 L/min. The triple quadrupole mass spectrometer (Ultivo, Agilent Technologies) was set to scan for transitions of individual peptides within scheduled 3-min windows around their retention time. Quadrupoles 1 and 3 were set to unit resolution. The acquired chromatograms were analyzed through Skyline 20.1.

Verification of BGLS biosynthetic genes in the BGLS$_g$ and BGLS$_p$ strains. The transformation was redone to generate the engineered strains. The same procedure was performed to harvest cultures from 24 h for analysis of targeted proteomics. Yeast cells collected from 200 μL culture were resuspended with 20 μL 10 mM NaOH. The cell solution was heated up at 95°C for 5 min, and 1 μL of the solution was used as the template for DNA amplification. The primers were designed to cover each whole gene and are listed in Table 3.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.7 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

This work was supported by the Danish National Research Foundation (DNRF99) and the Novo Nordisk Foundation (NNF17OC0027710). Sidsel Ettrup Clemmensen is thanked for helping to clone the construct for expressing *MET3* and *MET14*.

Conceptualization: C.W., C.C., and B.A.H. Experimental design: C.W., C.C., C.S.N., U.H.M., and B.A.H. Experimental implementation: C.W. generated all the constructs; C.S.N. generated the genome integration yeast strain BGLS$_g$; C.W. performed transformation, culturing, and metabolite extraction; C.W. and MP carried out targeted proteomics. LC-MS

method development and analysis: C.C. Statistical analysis: CW and MP. Writing—original draft: C.W. and B.A.H. Writing—review and editing: C.W., MP, C.C., C.S.N., U.H.M., and B.A.H. Funding: B.A.H. Supervision: B.A.H. and C.C.

We declare no conflicts of interest.

REFERENCES

1. Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303–333. https://doi.org/10.1146/annurev.arplant.57.032905.105228.
2. Traka M. 2016. Health benefits of glucosinolates. Advances in Botanical Res 80:247–279. https://doi.org/10.1016/bs.abr.2016.06.004.
3. Petersen A, Crocoll C, Halkier BA. 2019. De novo production of benzyl glucosinolate in *Escherichia coli*. Metab Eng 54:24–34. https://doi.org/10.1016/j.ymben.2019.02.004.
4. Wang C, Crocoll C, Agerbirk N, Halkier BA. 2021. Engineering and optimization of the 2-phenylethylglucosinolate production in *Nicotiana benthamiana* by combining biosynthetic genes from *Barbarea vulgaris* and *Arabidopsis thaliana*. Plant J 106:978–992. https://doi.org/10.1111/tpj.15212.
5. Petersen A, Wang C, Crocoll C, Halkier BA. 2018. Biotechnological approaches in glucosinolate production. J Integr Plant Biol 60:1231–1248. https://doi.org/10.1111/jipb.12705.
6. Geu-Flores F, Nielsen MT, Nafisi M, Møldrup ME, Olsen CE, Motawia MS, Halkier BA. 2009. Glucosinolate engineering identifies a γ-glutamyl peptidase. Nat Chem Biol 5:575–577. https://doi.org/10.1038/nchembio.185.
7. Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, Halkier BA. 2012. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng 14:104–111. https://doi.org/10.1016/j.ymben.2012.01.006.
8. Yang H, Qin J, Wang X, Ei-Shora HM, Yu B. 2020. Production of plant-derived anticancer precursor glucoraphanin in chromosomally engineered *Escherichia coli*. Microbiol Res 238:126484. https://doi.org/10.1016/j.micres.2020.126484.
9. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. 2015. Complete biosynthesis of opioids in yeast. Science 349:1095–1100. https://doi.org/10.1126/science.aac9373.
10. Trenchard IJ, Smolke CD. 2015. Engineering strategies for the fermentative production of plant alkaloids in yeast. Metab Eng 30:96–104. https://doi.org/10.1016/j.ymben.2015.05.001.
11. Srinivasan P, Smolke CD. 2019. Engineering a microbial biosynthesis platform for de novo production of tropane alkaloids. Nat Commun 10:3634. https://doi.org/10.1038/s41467-019-11588-w.
12. Srinivasan P, Smolke CD. 2020. Biosynthesis of medicinal tropane alkaloids in yeast. Nature 585:614–619. https://doi.org/10.1038/s41586-020-2650-9.
13. Batth TS, Keasling JD, Petzold CJ. 2012. Targeted proteomics for metabolic pathway optimization, p 237–249. *In* Keller N, Turner G (ed), Fungal secondary metabolism. Springer, New York, NY.
14. Masselot M, de Robichon-Szulmajster H. 1975. Methionine biosynthesis in *Saccharomyces cerevisiae*. Mol Gen Genet 139:121–132. https://doi.org/10.1007/BF00264692.
15. Albertsen L, Chen Y, Bach LS, Rattleff S, Maury J, Brix S, Nielsen J, Mortensen UH. 2011. Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes. Appl Environ Microbiol 77:1033–1040. https://doi.org/10.1128/AEM.01361-10.
16. Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, Forster J, Nielsen J, Borodina I. 2014. EasyClone: method for iterative chromosomal integration of multiple genes *Saccharomyces cerevisiae*. FEMS Yeast Res 14:238–248. https://doi.org/10.1111/1567-1364.12118.
17. Strucko T, Buron LD, Jarczynska ZD, Nødvig CS, Mølgaard L, Halkier BA, Mortensen UH. 2017. CASCADE, a platform for controlled gene amplification for high, tunable and selection-free gene expression in yeast. Sci Rep 7:41431. https://doi.org/10.1038/srep41431.
18. Ro D-K, Ouellet M, Paradise EM, Burd H, Eng D, Paddon CJ, Newman JD, Keasling JD. 2008. Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. BMC Biotechnol 8:83. https://doi.org/10.1186/1472-6750-8-83.
19. Lian J, Mishra S, Zhao H. 2018. Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: new tools and their applications. Metab Eng 50:85–108. https://doi.org/10.1016/j.ymben.2018.04.011.
20. Dornfeld KJ, Livingston DM. 1992. Plasmid recombination in a rad52 mutant of *Saccharomyces cerevisiae*. Genetics 131:261–276. https://doi.org/10.1093/genetics/131.2.261.
21. Thomas BJ, Rothstein R. 1989. Elevated recombination rates in transcriptionally active DNA. Cell 56:619–630. https://doi.org/10.1016/0092-8674(89)90584-9.
22. Isogai S, Okahashi N, Asama R, Nakamura T, Hasunuma T, Matsuda F, Ishii J, Kondo A. 2021. Synthetic production of prenylated naringenins in yeast using promiscuous microbial prenyltransferases. Metab Eng Commun 12:e00169. https://doi.org/10.1016/j.mec.2021.e00169.
23. Møldrup ME, Geu-Flores F, Olsen CE, Halkier BA. 2011. Modulation of sulfur metabolism enables efficient glucosinolate engineering. BMC Biotechnol 11:12–18. https://doi.org/10.1186/1472-6750-11-12.
24. Russel M, Model P, Holmgren A. 1990. Thioredoxin or glutaredoxin in *Escherichia coli* is essential for sulfate reduction but not for deoxyribonucleotide synthesis. J Bacteriol 172:1923–1929. https://doi.org/10.1128/jb.172.4.1923-1929.1990.
25. Burkart MD, Izumi M, Chapman E, Lin C-H, Wong C-H. 2000. Regeneration of PAPS for the enzymatic synthesis of sulfated oligosaccharides. J Org Chem 65:5565–5574. https://doi.org/10.1021/jo000266o.
26. Alani E, Cao L, Kleckner N. 1987. A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. Genetics 116:541–545. https://doi.org/10.1534/genetics.112.541.test.
27. Nour-Eldin HH, Hansen BG, Nørholm MH, Jensen JK, Halkier BA. 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res 34:e122. https://doi.org/10.1093/nar/gkl635.
28. Gietz RD, Schiestl RH. 2007. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. Nat Protoc 2:1–4. https://doi.org/10.1038/nprot.2007.17.
29. Ravilious GE, Nguyen A, Francois JA, Jez JM. 2012. Structural basis and evolution of redox regulation in plant adenosine-5′-phosphosulfate kinase. Proc Natl Acad Sci U S A 109:309–314. https://doi.org/10.1073/pnas.1115772108.
30. Mirza N, Crocoll C, Olsen CE, Halkier BA. 2016. Engineering of methionine chain elongation part of glucoraphanin pathway in *E. coli*. Metab Eng 35:31–37. https://doi.org/10.1016/j.ymben.2015.09.012.
31. Docimo T, Reichelt M, Schneider B, Kai M, Kunert G, Gershenzon J, D'Auria JC. 2012. The first step in the biosynthesis of cocaine in *Erythroxylum coca*: the characterization of arginine and ornithine decarboxylases. Plant Mol Biol 78:599–615. https://doi.org/10.1007/s11103-012-9886-1.
32. Crocoll C, Halkier BA, Burow M. 2016. Analysis and quantification of glucosinolates. Curr Protoc Plant Biol 1:385–409. https://doi.org/10.1002/cppb.20027.
33. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26:966–968. https://doi.org/10.1093/bioinformatics/btq054.
34. Von Der Haar T. 2007. Optimized protein extraction for quantitative proteomics of yeasts. PLoS One 2:e1078. https://doi.org/10.1371/journal.pone.0001078.
35. Wessel D, Flügge U. 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 138:141–143. https://doi.org/10.1016/0003-2697(84)90782-6.
36. Percy AJ, Chambers AG, Yang J, Domanski D, Borchers CH. 2012. Comparison of standard-and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. Anal Bioanal Chem 404:1089–1101. https://doi.org/10.1007/s00216-012-6010-y.
37. Batth TS, Singh P, Ramakrishnan VR, Sousa MML, Chan LJG, Tran HM, Luning EG, Pan EHY, Vuu KM, Keasling JD, Adams PD, Petzold CJ. 2014. A targeted proteomics toolkit for high-throughput absolute quantification of *Escherichia coli* proteins. Metab Eng 26:48–56. https://doi.org/10.1016/j.ymben.2014.08.004.